Proposed On-Chip Test Structure to Quantify Trap Densities within Flash Memories

Vandana Verma
Sr. Technical Marketing Engineer
Intel Corporation

Andrew Swaneck
Flash Product Engineer
Intel Corporation

Abstract

Degradation of the program/erase characteristics of Flash memory due to cycling is an industry wide reliability concern. This degradation in performance is associated with trapped charges present within the memory cells dielectric. The implementation of an on-chip test structure is proposed, allowing trapping characteristics of the Flash memory cells to be monitored. This paper discusses the on-chip test structure, program/erase characteristics of Flash memories, and electron trap density measurements.

Flash Memory Overview

Flash memory can be characterized as lying between the traditional Erasable Programmable Read Only Memory (EPROM) and Electrically Erasable Programmable Read Only Memory (EEPROM) components. Flash memory has the bitwise programmability of both memory structures but differs in its eraseability. Flash memory is electrically erasable unlike EPROM, which can only be erased using strong UV light. The electrical erase characteristics of Flash memory differs from an EEPROM in that control during erase is at the block or sector level and not at the bit level.

Flash memory is nonvolatile, in-system updateable, electrically erasable and is available in a variety of densities and footprints. The different component architecture of Flash are bulk erase, sector erase and boot block. Bulk erase components are those components in which the entire array is erased. Sector erase components are those components that are erased a sector at a time. A Flash block or sector being defined as a group of Flash cells and typically involves 512K bits. Boot block memory components offer flexibility in the size of the erase blocks. Additionally block locking capability allows enhanced code protection in one sector of the boot block memory components

Flash Program / Erase Overview

Flash memory can be derived from an EPROM base or an EEPROM base. The array structure can be classified as a NOR array or a NAND array. NOR Flash memory derives from an EPROM base. Figure-1 shows a typical cross sectional view an ETOX ™ Flash memory cell.

Proceedings of the 1996 IEEE International Workshop on Memory Technology, Design and Testing (MTDT '96)
1087-4852/96 $10.00 © 1996 IEEE

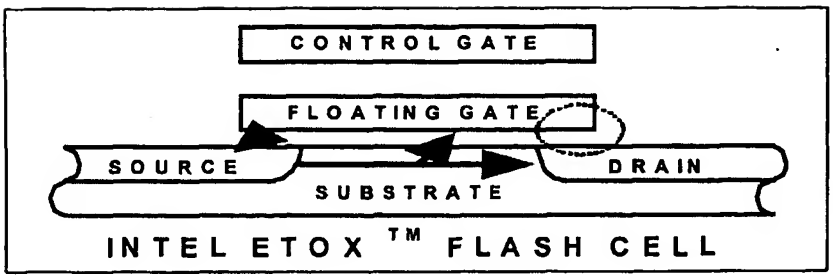


Figure 1. Intel ETOX Flash memory cell [1]

Physically, both programmability and erase involve transferring electrons to and from the "floating gate" of the Flash device, modifying the electrical field characteristics of the memory cell.

Flash Programming

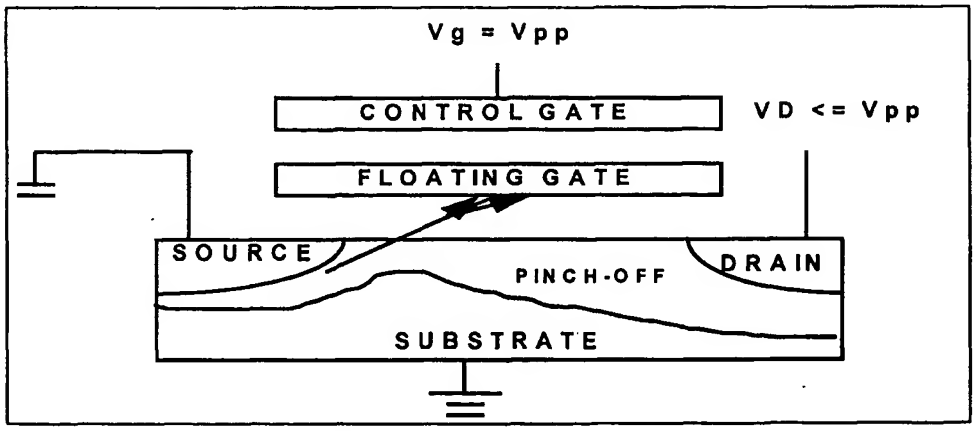
Changing the Flash memory cells electrical characteristics by storing charge on the floating gate is called programming. Programming a Flash cell involves applying 7.0V to the drain, 12.0V to the gate, and typically grounding the source of the device. The electric field established between the source and drain accelerate electrons towards the drain. Some electrons gain sufficient energy and are scattered within the channel. Some electrons are drawn towards the gate by the electric field established between the gate and the substrate. This process is known as "hot electron" programming. When reading the Flash cell, electrons present on the floating gate counteract the voltage on the select gate and prevent the memory cell from turning on. Since no current flows from drain to source, this is interpreted by the device as a zero on the memory component's output pin. Figure 2 shows an ETOX Flash cell being programmed.

Flash Erase

The mechanism which by which stored electric charge is removed from the floating gate is called erase. This is accomplished by applying 12.0V to the source and grounding the gate. There is a strong electric field established between the floating gate and the source. This electric field along with the thin oxide thickness, allow electrons with sufficient energy to pass through the oxide barrier. The mechanism by which a Flash cell is erased is called tunneling. Figure 3 shows a block diagram of a Flash memory cell during erase.

Flash Cycling

Cycling is defined as the number of times a Flash memory device can be erased and programmed in a reasonable amount of time without loss of device functionality at a specified failure rate percentage[1].



Figur 2. Intel ETOX Flash memory cell during programming

Proceedings of the 1996 IEEE International Workshop on Memory Technology, Design and Testing (MTDT '96)
1087-4852/96 $10.00 © 1996 IEEE

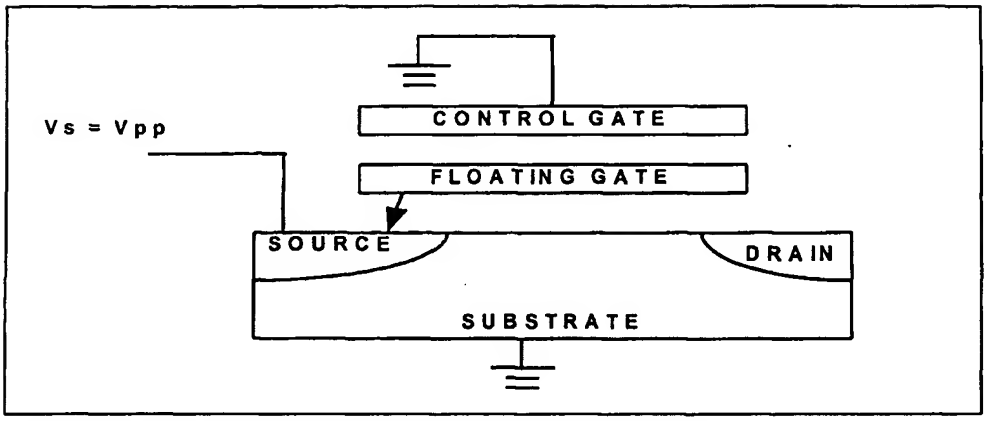


Figure 3. Intel ETOX Flash memory cell during programming

As the cycling number increases, electrons get trapped in the oxide region of the Flash memory cell resulting in increased program and erase times. This is however a non-destructive phenomenon. Cycling induced hard failures do not occur until an intrinsic wearout mechanism, caused by electron trapping, becomes so severe that program/ erase pulses can not overcome their opposing potential. The probability of this failure can be reduced and prolonged by using low electric fields, advanced low defect oxides and minimal oxide area per cell [2]. All steps should be taken to minimize cycling. Figure 4 shows the result of the degradation in programming performance due to cycling.

First Order Trapping Equation

Charge trapping within the dialectic modifies the potential barrier of the dielectric. This effectively increases the electric field necessary to achieve a given amount of charge injection through the dielectric. This distortion or charge trapping can be monitored by measuring the change in gate voltage necessary to maintain a constant current through the dielectric. Experimentally, a MOS structure was selected to study the feasibility of using constant current stressing, to determine the trapping characteristics within the dielectric. A high electric field was applied to the device structure, biasing the device into the tunneling current conduction region. Over time, and as electrons are trapped, the electric field necessary to maintain this constant tunneling current increases. This increase in electric field can be monitored as a change in applied gate voltage and is given by the following equation [4]

$$\mathrm{Ln}\left[dV_g / dF \right] = -\sigma F + \mathrm{Ln}\left[qXN_t\sigma / \varepsilon_{ox} \right] \qquad (1)$$

where F is the fluence or number of charge carriers through the dielectric, Nt is the density of charge traps, σ measures the ability of the trap site to capture a carrier (trap cross section), q is the electronic charge, ε_{ox} is the oxide permittivity, and X is the centroid of the trapped charge. Figure 5 shows a linear plot of the change in gate voltage measurements versus fluence, the slope of which is equal to the trap capture cross section. After the capture cross section has been determined, the density of the trap can be calculated from the plots intercept at zero fluence. The centroid of the trap charges is determined from the forward and reverse current-voltage characteristics before and after stressing. The centriod is given by the expression shown below [4]

$$X = T_{ox}\left[\Delta V_g^{+} / (\Delta V_g^{+} - \Delta V_g^{-}) \right] \qquad (2)$$

On Chip Trapping Test Structure

In order to utilize the first order trapping model and combine it with the cycling characteristics of a Flash cell, a on-chip test structure is proposed. This test structure is shown in Figure 6. The test structure consists of a Flash cell, modified such that the control gate is directly tied to the floating gate. This modification allows tight control of the gate voltage and the electric field across the oxide. Additionally, the current flowing through the oxide can be monitored. The addition of device test modes allow the voltages applied to the control gate, source, and drain of the test structure to be brought out to device pins. The constant current stressing technique described above can now be used to approximate trap densities within each Flash device. By comparing the voltage push-out seen during Flash cell cycling, to the trap density calculations performed on the test structure, a measure of the oxide quality can be

Proceedings of the 1996 IEEE International Workshop on Memory Technology, Design and Testing (MTDT '96)
1087-4852/96 $10.00 © 1996 IEEE

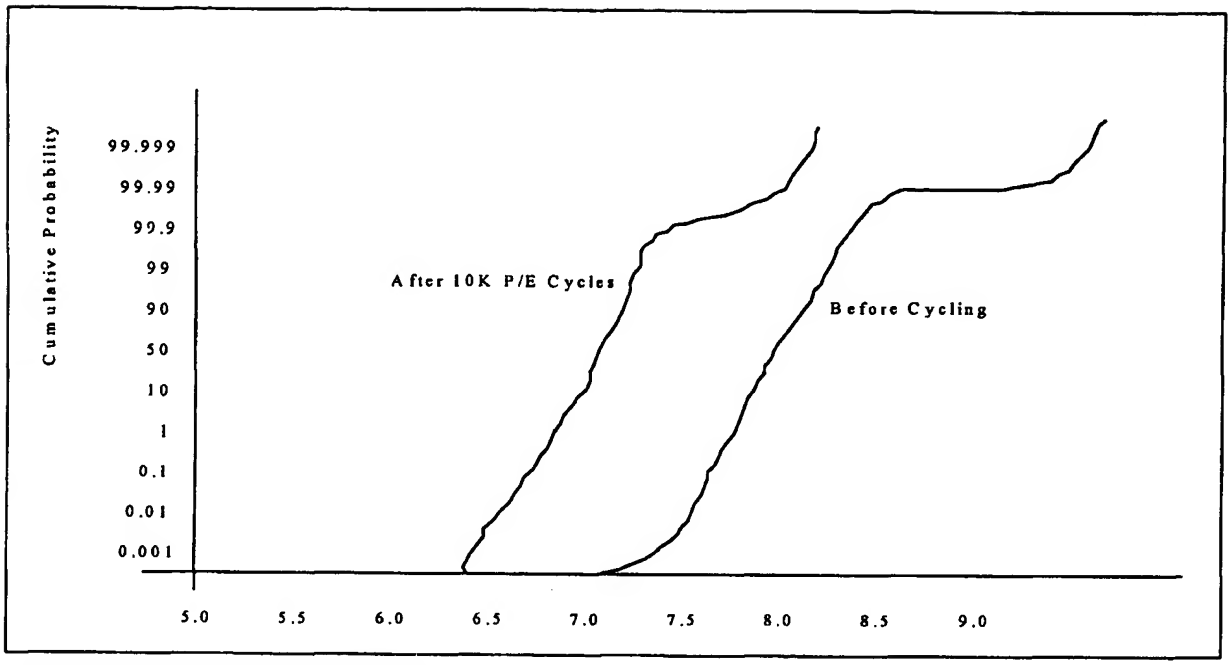


Figure 4. Program Vt vs. Cycles [3]

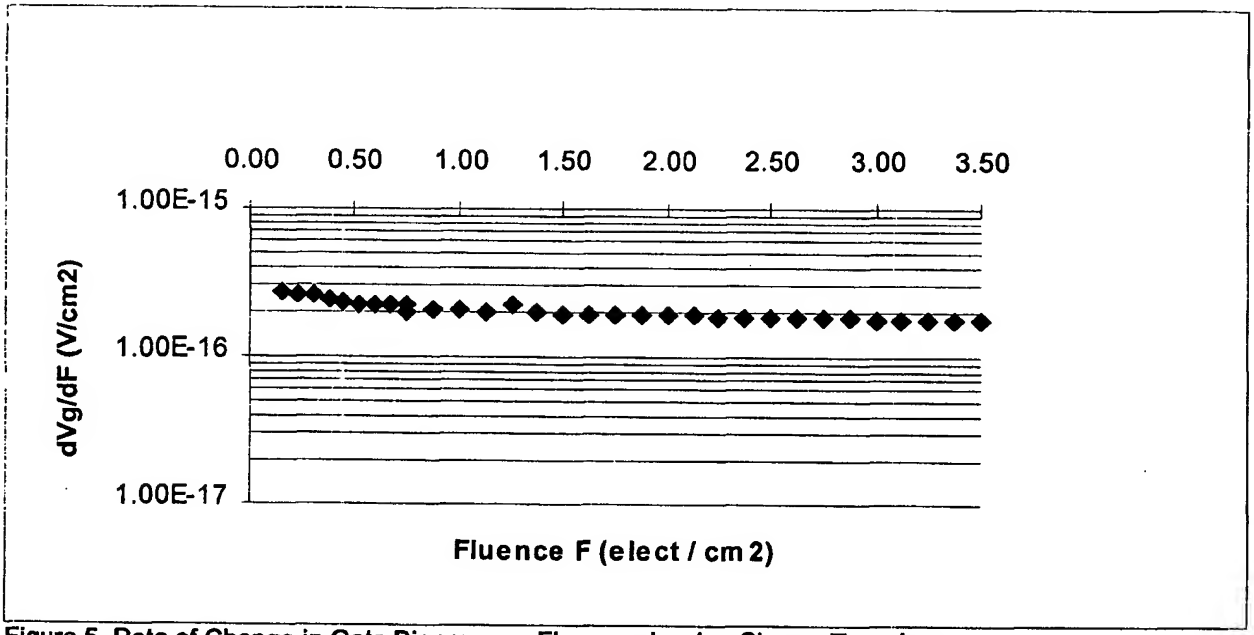


Figure 5. Rate of Change in Gate Bias versus Fluence showing Charge Trapping.

Proceedings of the 1996 IEEE International Workshop on Memory Technology, Design and Testing (MTDT '96)
1087-4852/96 $10.00 © 1996 IEEE

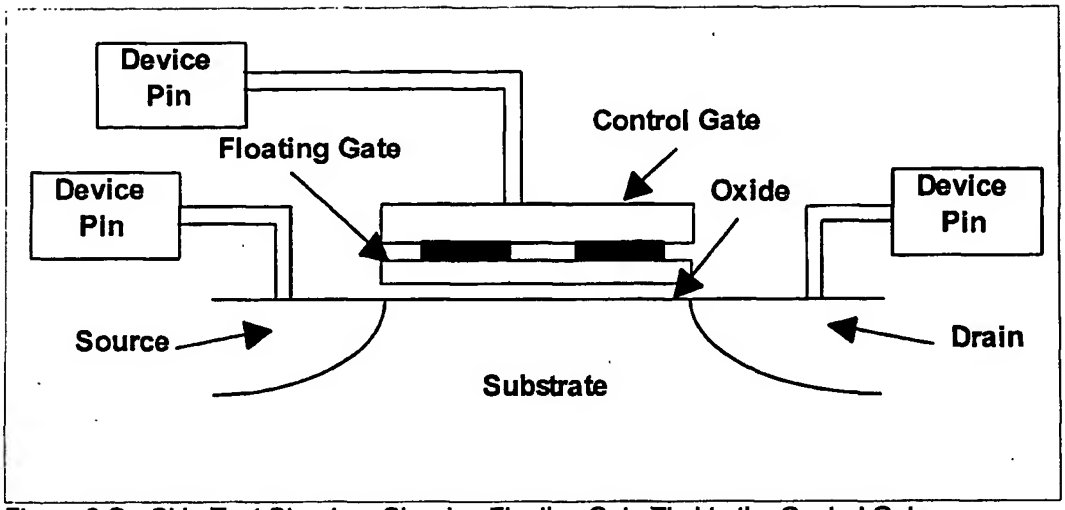


Figure 6 On-Chip Test Structure Showing Floating Gate Tied to the Control Gate

determined. Additionally, this structure enhances the ability to perform failure analysis upon cycling rejects and customer field returns. Subtle abnormalities in terms

of contamination or variations in processing conditions can be discovered. The increase in cost in terms of increased die size due to the addition of the test structure and test mode circuitry is minimal.

Conclusion

This paper discusses the importance of charge trapping and presents the degradation in performance associated with trapped charges. The proposed on-chip test structure allows the monitoring of the device's trapping characteristics. The program/erase characteristics of Flash memories, and electron trap density measurements were also presented.

References

[1] Brian Dipert, and Marcus Levy, "Designing with Flash Memory", Anna Books p38, p 143, 1993.
[2] Alan Buchecker, Jerry Kreifels, "The Intel 28F008SA Flash Memory,"p 18, 1993.
[3] M.S. Liang, C. Chang, Y.T. Yeow, C. Hu and R.W. Broderson, MOSFEST Degradation Due to Stressing of Thin Oxides," IEEE Trans. Electron Devcies, vol. ED-31, p 1238, 1984.

Proceedings of the 1996 IEEE International Workshop on Memory Technology, Design and Testing (MTDT '96)
1087-4852/96 $10.00 © 1996 IEEE